THE ROBUST DETECTION OF THE SEISMOACOUSTIC EMISSION SOURCES IN THE C-OTDR MONITORING SYSTEMS

Andrey Timofeev, Victor Denisov

Abstract. A method is proposed for stable detection of seismoacoustic sources in C-OTDR systems that guarantee given upper bounds for probabilities of type I and type II errors. Properties of the proposed method are rigorously proved. The results of practical applications of the proposed method in a real C-OTDR-system are presented in this report.

1 Introduction

Application of the C-OTDR (Coherent Optical Time Domain Reflectometer) technology to solve complex problems of remote monitoring of extended objects is currently being evaluated as a very promising approach. In particular, this technology can be effectively used to monitor oil and gas pipelines, controlling technological processes and identifying unauthorized activities in close proximity of the monitored objects. The core of the C-OTDR technology is the principle of highest vibrosensitivity of the single-mode optical fiber carrying a coherent infrared energy flow, which is injected into the fiber using a pulsed semiconductor laser. The main item of the C-OTDR technology is a comprehensive analysis of Rayleigh backscattered radiation characteristics, which transforms into an energetically weakened pulse and propagates constantly in the direction opposite to the direction of a pulsed laser flow. The reflected signal is created by the presence of static impurities in the optical fiber body and defects in the microstructure. The reflection is realized by a well-studied mechanism of elastic scattering of a Rayleigh type. There are thousands of Rayleigh scattering centers per one millimeter of fiber. Signals scattered by the centers coherently and randomly interfere with each other, forming so-called speckle patterns. Speckle patterns corresponding to different sections of the optical fibers are recorded and accumulated in the data center. The slightest change of the reflectance index value of the fiber, which occurred in a particular place, radically changes the speckle pattern corresponding exactly to this place of the fiber. These changes are reliably detected by the data center. The local changes in refractive index occur under the impact of temperature or due to mechanical action on the optical fiber surface. Let us call the optical fiber buried in the soil to a depth of 50-100 cm, an optical sensor (OS). Mechanical stress on the OS surface is caused by seismic acoustic waves. These waves are generated of the sources of elastic vibrations (SEV) which located in vicinity of laying optical fiber. Upon reaching the OS, seismoacoustic wave causes a local longitudinal microstrain on the surface of OS. Those microstrains in turn, cause a change in the local refractive index of light in a relatively small sector of the OS. As a result, the speckle pattern, which corresponds to this sector, changes significantly. Thus, the OS quite accurately reflects the state of the seismoacoustic field in its vicinity. The seismoacoustic field contains information about events that occur in the surface layers of the ground near the OS. This field is created by structural waves, which generated due to mechanical effects on the soil or as a result of a seismic activity. Walking or running man, traffic, earthworks, including hand digging are typical sources of the acoustic emission (structural acoustic wave).The SEV, which are subjects of interest for remote C-OTDR monitoring will be called a target SEV. For the convenience of data processing, the entire OS length is broken to successive portions (sites) each has length around 10-15 m. The data from those sites is processed separately. These sites will be called C-OTDR channels or just channels. Comprehensive analysis of the dynamic changes of speckle structures in each channel allows us to determine the type of SEV, which was detected in the channel. Determining the type of SAE is one of the main challenges for C-OTDR systems. However, in any case, the detection stage preceded the classification stage. At the stage of detection the SEV should provide practically reasonable values the upper bounds for errors of the first kind and the errors of second kind. The values of these boundaries are selected based on the requirements for effectiveness of the monitoring process, which should guarantee the necessary reliability when detecting SAE.

Described in the report the SAE detection algorithm guarantees the upper bounds for the probability of errors of the first and the second kinds. The proposed algorithm has been successfully tested in real conditions of the functioning of C-OTDR system, which was designed to monitor the ballast of railway tracks.

2 Statement of the problem

Indexes of channels C-OTDR system in conjunction form a set Z={1,2,…}. In every channel $j \in \mathbf{Z}$ is observed the intensity of the speckle-structure $S_j(t)$ as function of the time. Every channel represents the sector of OS of length N meters. Coordinates of every channel $j \in \mathbf{Z}$ are fully defined by the coordinates of its beginning $b(j)$ and ending $e(j)$ points. The 2-tuple $\left(b(j), e(j)\right)$ is called boundaries (start and end) of the channel j. If two channels match at least one of the boundaries, these channels are called adjacent. A group of adjacent channels (GAC) is the set of channels, each of which is adjacent to at least one of the GAC. The boundaries of all channels are given and form the tuple $\mathbf{X}(\mathbf{Z}) = \langle x_1, x_2, ... \rangle$. For j-th channel we have $\left(b(j), e(j)\right) = \left(\langle \mathbf{X}(\mathbf{Z}) \rangle_j, \langle \mathbf{X}(\mathbf{Z}) \rangle_{j+1}\right)$, here $\langle \mathbf{X} \rangle_j$ is j-th component of the tuple $\mathbf{X}(\mathbf{Z})$. Seismoacoustic wave propagating from SAE, reaches a certain GAC with delays those are proportional distance from the location of the SAE to a particular channel of GAC. The specific composition of GAC determined by the location SAE, its energetic power, distance of up to OS, as well as the parameters of the wave propagation medium. We call the GAC, which turned out under the influence by the seismoacoustic waves from SAE, the detecting GAC or DGAC.

Observations are made at successive times, which form a set $T = \{t_1, t_2, ...\}, \forall i > 0 : t_{i+1} - t_i = \Delta t > 0$. Thus, the observations are form the following sets $\mathbf{S}_j = \{S_j(t) | t \in T\}$, $j \in DGAC$. All channel and inter-channel statistics are defined on the intervals of duration Δ . Every of those intervals contain z of the discrete observations.

Specificity of the modern C-OTDR monitoring systems is such that any two channels are statistically independent only if no SAE, the elastic vibration from which affects the speckle patterns of these channels simultaneously. If such the SAE exist, those channels are become statistically dependent. In this case we can speak about a group of SAE, seismoacoustic waves from which will affect the OS channels simultaneously and compositionally. Obviously, the observations of any two channels of the DGAC are statistically dependent. We call the area the sensitivity of the C-OTDR system the area Ω which situated in vicinity of OS and with the appearance inside which one or a group of SAE, the observations $\mathbf{S}_j, j \in DGAC,$ of the appropriate DGAC will be abruptly change their statistical characteristics. In other words, observations $\mathbf{S}_j, j \in DGAC$ will are mutually dependent after appearance insideΩone or a group of SAE. We denoted:

- τ is the moment of abrupt change of the observations distributions, which happened because of appearance the SAE in Ω; actually, τ is a point of appearance the signals from the SAE in DGAC or **change-point moment**;
- hypothesis H_0 : in the region Ω do not SAE (background model);
- hypothesis H_1 : in the region Ω is at least one SAE (signaling model);
- $\alpha \in]0,1[$ is a predetermined upper bound for the probability of making type I errors;
- $\beta \in]0,1[$ - is a predetermined upper bound for the probability of making type II errors;
- $\mathbf{\Delta}(t) = [t - z, t]$ is the interval for calculation of the speckle-structures statistical characteristics (speckle-metrics), where z is the number of the discrete observations inside of the interval $\mathbf{\Delta}(t)$; during monitoring the interval $\mathbf{\Delta}(t)$ is shifted by an amount Δt along the time axis T;
- $\widetilde{S}_j(t, t+z) = \{S_j(t), S_j(t+1), ..., S_j(t+z)\} \subseteq \mathbf{S}_j$;
- $\overline{\widetilde{S}}_j(t_1, t_2) = \sum_{t=t_1}^{t_2} S_j(t)(t_2 - t_1)^{-1}, \overline{\overline{S}}_j(t_1, t_2) = \sum_{t=t_1}^{t_2} \left(S_j(t) - \overline{\widetilde{S}}_j(t_1, t_2)\right)^2$;

- $r^{(i,j)}\left(t_1,t_1+z\mid\delta\right)=\dfrac{\sum_{t=t_1}^{t_1+z}\left(\widetilde{S}_j(t)-\overline{\widetilde{S}}_j(t_1,t_1+z)\right)\left(\widetilde{S}_i(t+\delta)-\overline{\widetilde{S}}_i(t_1+\delta,t_1+z+\delta)\right)}{\left(\overline{\overline{S}}_j(t_1,t_1+z)\overline{\overline{S}}_i(t_1+\delta,t_1+z+\delta)\right)^{0.5}}$;
- $\mathbf{\Sigma}(\mathbf{Z})=\bigcup_{j\in DGAC}\mathbf{S}_j$;

- $\rho\left(t\mid\mathbf{\Sigma}(\mathbf{Z}),\mathbf{\Delta}(t)\right)\subseteq R^1$ is some stochastic function, which we will call the signaling function; this function is defined on the interval $\mathbf{\Delta}(t)$ and depends on the $\mathbf{\Sigma}(\mathbf{Z})$ so that: $\mathbf{E}\left(\rho\left(t\mid\cdot\right)\middle|H_0\right)=0$, $\mathbf{E}\left(\rho\left(t\mid\cdot\right)\middle|H_1\right)>0$, $\mathbf{E}\left(\rho^2\left(t\mid\cdot\right)\middle|H_0\right)<\infty$.

Watching $\mathbf{S}_j=\left\{S_j(t)\middle|t\in T\right\},j\in\mathbf{Z}$, need to define a function $W\left(t,z,\rho,\alpha,\beta\right)\in\left\{0,1\right\}$, that depends on the $\rho\left(\cdot\right)$, $\mathbf{\Delta}(t)=\left[t-z,t\right]$ and on the parameters α,β such that

- $\mathrm{P}\left[W(t,z,\rho,\alpha,\beta)=1\middle|\tau\notin\left[t-z,t\right]\right]\le\beta$,
- $\mathrm{P}\left[W(t,z,\rho,\alpha,\beta)=0\middle|\tau\in\left[t-z,t\right]\right]\le\alpha$.

Thus, it has been tasked interval estimation of the change-point τ (τ is the moment of the appearance the SAE in Ω).The solution should guarantee the predetermined upper bounds for the probabilities of making type I(α) and type II (β) errors. In this formulation, the problem of detection SAE reduces to problem of interval estimation of change-point τ.

3 Selecting a robust signaling function ρ

By definition the signaling function $\rho\left(t\mid\mathbf{\Sigma}(\mathbf{Z}),\mathbf{\Delta}(t)\right)$ abruptly changes its average value (towards increase) at the moment τ. It's desirable to the probability distribution of the ρ had been robust to outliers in the observations $\mathbf{\Sigma}(\mathbf{Z})$. By the condition of the problem statement, vectors $\widetilde{S}_j\left(t,t+z\right)$, $\widetilde{S}_i\left(t+\delta,t+z+\delta\right)$ satisfy the conditions of Theorem 8.1 [2], p. 204, if $t+z+\delta<\tau$. Following this theorem, $\forall_{i,j,\delta}\left\{\mathbf{E}\left(r^{(i,j)}\left(t,t+z\mid\delta\right)\right)=0,\ \mathbf{E}\left(r^{(i,j)}\left(t,t+z\mid\delta\right)\right)^2=\left(z-1\right)^{-1}\right\}$. As the problem statement dictates we can use the following function as a signaling one:

$$\rho\left(t\mid\mathbf{\Sigma}(\mathbf{Z}),\mathbf{\Delta}(t)\right)=\sup_{i,j\in DGAC}\left[\sup_{\delta}\left(r^{(i,j)}\left(t,t+z\mid\delta\right)\right)\right]. \quad (1)$$

However, this function is not robust to the statistical anomalies of observations. The robust estimates of the correlation coefficient were considered in [1]-[4]. Specifically, in section 8.3 [2] a powerful approach was suggested to obtain the robust estimates. Following [2], when calculating $r^{(i,j)}(t,t+z\mid\delta)$, instead $\widetilde{S}_j(\cdot),\widetilde{S}_i(\cdot)$ uses some $\mathbf{u}\left(\widetilde{S}_j(\cdot)\right),\mathbf{v}\left(\widetilde{S}_i(\cdot)\right)$, which were calculated from $\widetilde{S}_j(\cdot),\widetilde{S}_i(\cdot)$ respectively, according to next five rules: (1) $\mathbf{u}=\Psi\left(\widetilde{S}_j(\cdot)\right),\mathbf{v}=\Xi\left(\widetilde{S}_i(\cdot)\right)$; (2) Ψ,Ξ commute with permutations of the components of $\widetilde{S}_j(\cdot),\mathbf{u}$ and of $\widetilde{S}_i(\cdot),\mathbf{v}$; (3) Ψ,Ξ preserve a monotone ordering of the components of $\widetilde{S}_j(\cdot)$, $\widetilde{S}_i(\cdot)$; (4) $\Psi=\Xi$; (5) $\forall a>0,\forall b,\exists a_1>0,\forall b_1,\forall\mathbf{x}\Psi(a\mathbf{x}+b)=a_1\Psi(\mathbf{x})+b_1$. In the following two examples [2], all five requirements hold:

- the classical Spearman rank correlation between $\widetilde{S}_j(\cdot)$ and $\widetilde{S}_i(\cdot)$;
- the quadrant correlation between $\widetilde{S}_j(\cdot)$ and $\widetilde{S}_i(\cdot)$.

If the $r^{(i,j)}(t,t+z\mid\delta)$ was calculated by means of the classical Spearman rank, we will denote this as $r_{sr}^{(i,j)}(t,t+z\mid\delta)$. And if the $r^{(i,j)}(t,t+z\mid\delta)$ was calculated using the quadrant correlation, we will use the following notation $r_{qc}^{(i,j)}(t,t+z\mid\delta)$. Both of them $r_{sr}^{(i,j)}(t,t+z\mid\delta)$ and $r_{qc}^{(i,j)}(t,t+z\mid\delta)$ are robust to the statistical anomalies of observations. Therefore, the corresponding signaling functions

$$\rho_{sr}(t\mid\mathbf{\Sigma}(\mathbf{Z}),\mathbf{\Delta}(t))=\sup_{i,j\in DGAC}\left[\sup_{\delta}\left(r_{sr}^{(i,j)}(t,t+z\mid\delta)\right)\right]$$

and

$$\rho_{qc}(t\mid\mathbf{\Sigma}(\mathbf{Z}),\mathbf{\Delta}(t))=\sup_{i,j\in DGAC}\left[\sup_{\delta}\left(r_{qc}^{(i,j)}(t,t+z\mid\delta)\right)\right]$$

are robust too. It easy to see:

$$\forall t<\tau:\left[\mathbf{E}\rho_{sr}(t\mid\mathbf{\Sigma}(\mathbf{Z}),\mathbf{\Delta}(t))=0,\mathbf{E}\rho^2_{sr}(t\mid\mathbf{\Sigma}(\mathbf{Z}),\mathbf{\Delta}(t))=1/(z-1)\right], \tag{2}$$

$$\forall t<\tau:\left[\mathbf{E}\rho_{qc}(t\mid\mathbf{\Sigma}(\mathbf{Z}),\mathbf{\Delta}(t))=0,\mathbf{E}\rho^2_{qc}(t\mid\mathbf{\Sigma}(\mathbf{Z}),\mathbf{\Delta}(t))=1/(z-1)\right].$$

Thus, functions ρ_{sr} and ρ_{qc} may be used as the robust signaling functions (RSF).The RSF we will denote as $\rho_{rb}\in\{\rho_{sr},\rho_{qc}\}$.

4 Guaranteed detection method of the SAE

As it follows from (2), until the moment τ the expectation of the RSF is zero. Under influence of elastic vibrations from SAE, expectation of the RSF abruptly changes towards increase, because the observations at once several channels of DGAC become statistically dependent. In other words, at the moment τ the model of RSF will get change-point of its probabilistic properties compared with the background model H_0. Let us describe the method for interval estimation of the moment τ.

Remark 1*. If* $W(t,z,\rho,\alpha,\beta)=1$ *the* $[t-z,t]$ *will be confidence interval for moment* τ*. In the case of realization the event* $W(t,z,\rho,\alpha,\beta)=1$ *decision will be taken that* $\tau\in[t-z,t]$*. Essentially - it is a fact of detection SAE. And it is guaranteed the predetermined upper bounds for the probabilities of making type I (α) and type II (β) errors.*

Properties of the proposed method are described in the following theorem:

Theorem 1. Let

1. $\exists C>1:\mathbf{Var}\left(\rho_{rb}\left(t\mid\mathbf{\Sigma}(\mathbf{Z}),\mathbf{\Delta}(t)\right)\right)\le Cz^{-1}$;
2. $\exists\theta>0:\inf_{t\ge\tau}\mathbf{E}\rho_{rb}\left(t\mid\mathbf{\Sigma}(\mathbf{Z}),\mathbf{\Delta}(t)\right)\ge\theta$;
3. $W(t,z,\rho,\alpha,\beta)=\begin{cases}1, if\ \rho_{rb}\left(t\mid\mathbf{\Sigma}(\mathbf{Z}),\mathbf{\Delta}(t)\right)/\theta\ge b\\0, if\ \rho_{rb}\left(t\mid\mathbf{\Sigma}(\mathbf{Z}),\mathbf{\Delta}(t)\right)/\theta<b\end{cases}$*, here*

$$b=\left(\frac{\alpha}{\beta C^2}\right)^{0.5}\left(1-\left(\frac{\alpha}{\beta C^2}\right)^{0.5}\right)^{-1},$$

Then, if $z=1+\frac{C^2}{\theta^2\alpha}-\frac{2C}{\theta^2\sqrt{\alpha\beta}}+\frac{1}{\theta^2\beta}$, the following statements are true:

$$\mathrm{P}\left[W(t,z,\rho,\alpha,\beta)=1\middle|\tau\notin[t-z,t]\right]\le\beta$$
$$\mathrm{P}\left[W(t,z,\rho,\alpha,\beta)=0\middle|\tau\in[t-z,t]\right]\le\alpha$$

Remark 2. *The constant θ has following sense: the algorithm will detect those SAE, for which the increment of the value* $\mathbf{E}\rho_{rb}\left(t \mid \boldsymbol{\Sigma}(\mathbf{Z}), \boldsymbol{\Delta}(t)\right)$ *when* $t - z \geq \tau$ *will exceed the value* θ. *Thus, parameter* θ *defines the sensitivities of the detection algorithm.*

Proof of Theorem 1. Consider the representations

$$\rho_{rb}\left(t \mid \boldsymbol{\Sigma}(\mathbf{Z}), \boldsymbol{\Delta}(t)\right)/\theta = \begin{cases} m(t), t < \tau \\ \mathbf{E}\rho_{rb}\left(t \mid \boldsymbol{\Sigma}(\mathbf{Z}), \boldsymbol{\Delta}(t)\right)/\theta + m'(t), t - z \geq \tau \end{cases}$$

In view of (2) we have:

$$\mathbf{E}m(t) = 0, \mathbf{E}m^2(t) = 1/(\theta^2(z-1)). \tag{3}$$

Using Chebyshev's inequality, we write:

$$\begin{aligned} &\mathrm{P}\left[W(t,z,\rho,\alpha,\beta) = 1 \middle| \tau \notin [t-z,t]\right] = \\ &\mathrm{P}\left[\rho_{rb}\left(t \mid \boldsymbol{\Sigma}(\mathbf{Z}), \boldsymbol{\Delta}(t)\right)/\theta > b \middle| \tau \notin [t-z,t]\right] \leq \\ &\mathrm{P}\left[|m(t)| > b\right] \leq (\theta^2(z-1)b^2)^{-1} \end{aligned} \tag{4}$$

When $\tau > t - z$ from condition 2 we have $\mathbf{E}\rho_{rb}\left(t \mid \boldsymbol{\Sigma}(\mathbf{Z}), \boldsymbol{\Delta}(t)\right)/\theta \geq 1$. Therefore, taking into account the first condition of the theorem, we can write:

$$\begin{aligned} &\mathrm{P}\left[W(t,z,\rho,\alpha,\beta) = 0 \middle| \tau \in [t-z,t]\right] = \\ &\mathrm{P}\left[\mathbf{E}\rho_{rb}\left(t \mid \boldsymbol{\Sigma}(\mathbf{Z}), \boldsymbol{\Delta}(t)\right)/\theta + m'(t) < b \middle| \tau \in [t-z,t]\right] \leq \\ &\mathrm{P}\left[\left|\mathbf{E}\rho_{rb}\left(t \mid \boldsymbol{\Sigma}(\mathbf{Z}), \boldsymbol{\Delta}(t)\right)/\theta\right| - |m'(t)| < b \middle| \tau \in [t-z,t]\right] \leq \\ &\mathrm{P}\left[|m'(t)| > \left|\mathbf{E}\rho_{rb}\left(t \mid \boldsymbol{\Sigma}(\mathbf{Z}), \boldsymbol{\Delta}(t)\right)/\theta\right| - b\right] \leq \\ &\mathrm{P}\left[|m'(t)| > 1 - b\right] \leq C^2/(z\theta^2(1-b)^2) \leq C^2/((z-1)\theta^2(1-b)^2) \end{aligned} \tag{5}$$

Substituting in (4) and (5) the values of z and b, as defined in the condition of the theorem immediately confirm the truths of the allegations are proved. The theorem is proved.

5 Usage of the suggested approach in practice.

The approach described in this report is used for the detection of SEV in real C-OTDR system. Parameters of the C-OTDR system:

- the probe pulse duration - 10 ns;
- frequency sensing - 8 kHz,
- the probe signal power - 15 mW ,
- laser wavelength - 1550 nm .

Table 1 contains results of detection the SEV. Here «Distance» is an average distance at which the given class of SEV was detected, P_I - is detection error of the type I; P_{II} - error of the type II. Parameters of the detection system were such: z=25, $\theta = 0.2$, $\alpha = 0.1$, $\beta = 0.1$. The parameter C was estimated experimentally: C~1.3.

Data are presented for rocks cemented soil. The results in Table1 show sufficiently high practical effectiveness of the described SEV detection system.

Table 1. The practical detection results

Type of SEV	Distance (m)	P_I	P_{II}	Type of SEV	Distance (m)	P_I	P_{II}
"hand digging the soil"	10	0.09	0.08	"easy excavation equipment"	40	0.03	0.02
"chiselling ground scrap"	5	0.1	0.1	"spilling fluid pressure",	30	0.09	0.1
"walking man"	10	0.07	0.09	"liquid spills without pressure"	10	0.1	0.1
"running man"	15	0.09	0.08	"movement of the cleaning scraper"	10	0.03	0.02
"passenger car"	25	0.06	0.1	"cutting frozen soil"	15	0.02	0.04
"truck"	35	0.07	0.09	"shrew digging the ground"	20	0.06	0.07
"heavy equipment excavator"	50	0.02	0.01				

References

1. F. R Hampel, E.M. Ronchetti, P.J. Rousseeuw and W.A. Stahel. *Robust Statistics. The Approach Based on Influence Functions.* New York: John Wiley, 1986.
2. P. Huber, *Robust Statistics*, John Wiley & Sons, New York, 1981.
3. R. Maronna, D.M. Yohai, *Robust Statistics. Theory and Methods*. New York: John Wiley & Sons, 2006.
4. G. Shevlyakov, P. Smirnov, "Robust Estimation of the Correlation Coefficient: An Attempt of Survey", *Austrian Journal of Statistics*, Vol. 40, No. 1 & 2, pp.147–156, 2011.

Timofeev Andrey V. was born in Chita (Russia). He received Dr. habil. sc. ing. in Computer and Information Sciences from Tomsk State University of Control Systems and Radioelectronics, Russia, in 1994. A number of research publications in the International journals (JKSS, Stat.Methodology., Automation and Remote Control etc) and International/National conferences are at his credit. He is on the editorial board of several journals and conferences and a referee of several others. His research interests include non-asymptotic nonlinear methods of confidence estimation of multidimensional parameters of stochastic systems; machine learning, large margin classification in Banach spaces; confidence Lipschitz classifiers; technical diagnostics, C-OTDR systems; data mining; change-point problem; alpha-stable laws; statistical classification in application to biometrics and seismics.